Health in data space. Formative and Experiential Dimensions of Cross-Border Health Data Sharing

Hoeyer, Klaus; Green, Sara; Martani, Andrea; Middleton, Alexandra Suzanne; Pinel, Clémence Pauline Cécile

Published in:
Big Data & Society

DOI:
10.1177/20539517231224258

Publication date:
2024

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NG-ND

Citation for published version (APA):
Health in data space: Formative and experiential dimensions of cross-border health data sharing

Klaus Hoeyer1, Sara Green2, Andrea Martani3, Alexandra Middleton1 and Clémence Pinel1

Abstract
Healthcare is increasingly dataized, and a wide range of actors—patients, clinicians, administrators, policymakers, and industry lobbyists—want to be able to exchange and access health data internationally and use them for an increasing number of purposes. Therefore, competing initiatives aimed at fostering international data integration proliferate, with the proposed European Health Data Space as one of the most prominent examples. But how do legislators conceptualize a health data space? And what could they gain from rethinking the governmental object of this legislation? To explore these questions, we suggest taking the term “data space,” present in the European Health Data Space initiative, and develop it theoretically to establish a vocabulary fit for understanding international data-intensive health environments.

Space is a concept with appealing affordances. It is a way of naming a mode of being which is simultaneously symbolic and material, abstract and concrete, social and physical. We show how these affordances of the concept of space can be helpful when exploring new ways of living in cross-border data-intensive healthcare settings. Whereas policy reports often describe data sharing as a matter of providing technical means and legal provisions to “wire together” existing data resources, we argue that data spaces should be understood as sociotechnical constructs enacted through three formative and four experiential dimensions.

Keywords
Data infrastructure, data politics, data sharing, European health data space, EU, law

Introduction
Throughout the global health policy landscape, powerful actors seek to facilitate new ways of making health data available across domains, national borders, and purposes. How should this movement, and its implications for the everyday lived experience of healthcare, be understood? In this article, we propose a new conceptual and theoretical framework outlining a way to think about the data spaces currently being constructed. By examining the European Commission’s proposal for a European Health Data Space (EHDS), we explore what policymakers appear to envisage—and propose an alternative way of conceptualizing “health data space” as a governmental object. As we write this, the EHDS proposal is just that—a proposal. It will not be passed without substantial changes. However, this vision is worth examining because it informs and elucidates the discourses used by key policymakers in the European Union. Furthermore, it prompts us to contemplate how citizens and health professionals might experience such changes in their everyday lives. Indeed, the envisioned forms of governance may affect the organization of healthcare systems, with ramifications for health professionals and citizens throughout the European Union (EU).

We wish to mobilize a theoretical awareness of the object of governance that proposals like the EHDS call into being. On the surface, this initiative appears to be primarily concerned with designating legal entitlements and...
connecting wires to allow standardized data to cross domains and borders. It thus presents mainly as a legal and technical change that, once solved, will serve a range of already-existing desires: patient empowerment, greater knowledge, efficient governance, increased wealth, and European integration. However, digital integration is much more than a technical and legal issue (Hoeyer, 2023). It is social and political to its core (Bellanova and Duez, 2012). Initiatives like the EHDS ignite a set of broader transformations concerning the ways data permeate everyday life (Marelli et al., 2023) and, to grasp these transformations, we need a language capable of engaging with the forces that bring a “health data space” into being. By building on classic writing on space (Durkheim, 2008; Lefebvre, 1974), and key insights from digital geography (Kitchin and Dodge, 2011; Zook and Graham, 2007a, 2007b), this paper seeks to develop such a language.

Traditionally, health data have been produced primarily by health professionals and stored in local healthcare systems, and therefore have been more nationally bounded than many other types of data. They have thus far been intended primarily for clinical purposes. Even when used for research or administrative planning, these purposes have been framed as “secondary” in the literal sense (i.e. “subordinate”). This is no longer the case. Today, health providers increasingly collect data in formats designed to serve aims other than clinical care (Büchner, 2018; Hunt et al., 2017). This blurs the distinction otherwise typically used by clinicians between primary (for the sake of treatment) and secondary (for the sake of research, administration, and innovation) uses of health data. Furthermore, health data are no longer produced only by health professionals, but also by citizens using wearables. Tech companies continuously release new digital health technologies, online tests, and integrated online services capable of tracing citizens as healthcare consumers (Roberts et al., 2019). Meanwhile, policymakers around the globe have bought into the notion of health data as “assets” that ensure economic growth via biomedical research and innovation (Tarkkala et al., 2018; Vezyridis and Timmons, 2021), and with this in mind, they seek to build new data infrastructures.

In this context, and on the heels of the COVID-19 pandemic, the EHDS has been put forth as the first of a series of planned European “data spaces.” As mentioned above and detailed below, the European Commission envisages the EHDS as a technical infrastructure operating on a new legal mandate to make health data accessible for citizens, clinicians, governmental actors, and companies across the entire union. There is, however, more at play. Beyond just interconnecting wires, the social fabric of healthcare relations will be impacted by increased data sharing, with implications for how patients and people in the health services perceive both each other and healthcare itself. To uncover these layers—and thereby conceptualize more precisely the governmental object of initiatives such as the EHDS—we propose taking the concept of data space and developing it theoretically. This endeavor can help us think critically about the type of health services that people will experience when pursuing health in a time of increased digital connectedness. We purport that when properly operationalized, the concept of data space allows us to understand the workings of these new forms of data integration as operating in both digital and physical territories, and as enacted through both social and technical means.

Below, we develop our theoretical proposition. Equipped with this lens, we turn to the EHDS proposal and employ this proposition to analyze how policymakers describe the desired changes embedded in the EHDS initiative. We primarily analyze official documents, though we mention a few informal discussions with civil servants and other stakeholders at national and EU levels where we find them illuminating. Rather than presenting an empirical analysis of the ongoing processes around the proposal, we read it in light of previous work—our own and that of others—on the social implications of integration of health data infrastructures. Our aim remains conceptual rather than empirical.

Reappropriating the notion of data space

The notion of “data space” is already present in the very name of the EHDS initiative, but not in any theoretically developed way. In information science, “data space” was originally used to describe the shift from structured databases with high semantic integration to ways of searching for information in heterogeneous and volatile environments (Franklin et al., 2005). The emphasis at the time was on methods of information seeking. Now the Commission imbues the term with new potency when they write on the EHDS homepage:

The European Health Data Space is a health-specific ecosystem comprised of rules, common standards and practices, infrastructures and a governance framework that aims at empowering individuals through increased digital access to and control of their electronic personal health data, at national level and EU-wide, and support to their free movement, as well as fostering a genuine single market for electronic health record systems, relevant medical devices and high risk AI systems.

This description focuses on legal entitlements and technical interoperability. Our suggestion, however, is to take the term “data space” and reinvigorate it as a sociotechnical concept. This helps us better understand what proposals for increased data-interconnectedness like the EHDS might produce. Rather than being a neutral “container” of
health data, data spaces are simultaneously constituted by and constituting for human activities. They are enacted, so to speak. We suggest understanding data spaces as enacted through three formative elements (what we call promises, work, users) and experienced in at least four dimensions, relating to what people experience as right (legally and morally), true (epistemologically), present (phenomenologically), and valuable (economically, emotionally, and socially), graphically outlined in Table 1. Each of these dimensions may conflict with each other or intersect in various ways. We begin by explicating this understanding of data space before turning to an analysis of the EHDS proposal based on such understanding.

**Data space as a generative concept**

Data ubiquity reflects a form of cyborg life permeated with technology (Haraway, 2004). Indeed, the Internet of Things structures human movement in interconnected spaces via online data processing to such an extent that Floridi (2009) suggests “it is no longer sensible to ask whether one may be online or offline” (p. 1). Important work in digital geography has pointed out how software permeates social life by shaping options and relations. Kitchen and Dodge talk about code/space as those spaces that only function thanks to various forms of programming (Kitchin and Dodge, 2011). Zook and Graham (2007a) point to the ways in which such “new hybrid spaces (built through the dynamic melding of code, data and material place) affect how people live and the extent to which these spaces are understood and shaped in non-transparent ways” (p. 1331). The software through which data become agents remains invisible, as Kitchin and Dodge point out, and for most people it is largely incomprehensible. They experience not the programming, only the effects. Because data depend on physical infrastructures, data spaces should not be thought of as purely virtual objects. Parallel to Lessig’s (2006) argument that cyberspace, though built on the Internet, is a much richer experience than merely connected wires, we think it is important to explore data spaces as experiential objects. Data are exchanged in ways not necessarily understood from the vantage points of citizens, clinicians, administrators, or companies (Haddow et al., 2010). Still, the new availability of data redefines the social and technological conditions for patients to interact with the healthcare system. Data spaces are how materiality and meaning merge in data-saturated lives.

How does the concept of data space allow us to capture all this? Consider the affordances of the term space itself. A space designates an area as well as a time interval, a gap as well as a filling. Spaces designate frames for human activity, but they are simultaneously produced and shaped by humans as they experience and inhabit them in particular ways.° Space is at once bounded (a space) and without or beyond boundaries (the space). Space can also mark absence, emptiness, and between-ness (negative space). When space alludes to a form of infinity, as in an extraterrestrial space, it invites us to think about the unknown, that beyond our control or realm of perception. Data spaces—like extraterrestrial space—involve unknown potentials because data can always be linked in new ways (Bowker, 2005). These affordances for merging materiality and meaning, presence and absence, boundary and boundlessness are useful when contemplating the type of governmental object targeted by cross-national data infrastructures for health.

Indeed, the term space has been theorized in human geography and Marxist sociology precisely as a way of moving dialectically between the concrete and the abstract (Elden, 2009). As Durkheim (2008) once remarked: space is physically, symbolically, and socially constructed (Knoblauch and Martina, 2020). Lefebvre (1991) similarly argued that humans move in spaces that are simultaneously representations (symbols) and materiality (things), both imagined and experienced. This observation has new pertinence when exploring data worlds. Data are material—they need a medium to exist (O’Riordan, 2017)—and they exert agency as symbols (Thrift, 2008). We, therefore, suggest defining data spaces as conditions of being that are enacted when people, computers, and organizations use data for maneuvering hopes, opportunities, risks, and challenges—or think that others might be doing so.

The point with data spaces is that the virtual lives of data and the physical lives of patients and health professionals are not separable. Awareness of data increasingly influences physical interactions between patients and health professionals in all kinds of ways (Duckert and Barkhuus, 2022; Heinesen et al., 2022; Kristensen et al., 2021). This makes the traditional conceptual boundaries between virtual and real, digital and physical, linguistic and material, signifier and signified, misleading. Still, we should not assume that computers and bodies are simply “fusing” or becoming “the same” (Boellstorff, 2012). Databases speak their own language, one that patients often do not understand. Data acquire agency through software that remains invisible and even incomprehensible to most people (Kitchin and Dodge, 2011). Humans generally use and understand information differently from computers; just as computers cannot engage with or interpret patients in the same way as a doctor can. Therefore, the virtual and the actual are not fusing. Computers do not access human lived experience. Human experience, conversely, cannot clearly determine at which point data stop and the world begins, or vice versa. This is yet a reason for thinking about experiential dimensions (how people live with data) as different from the formative dimensions (the infrastructures that are built). By reappropriating the term data space, we wish to bring attention to
the bodily, phenomenological experience of moving in spaces where data are always silent actors. These spaces are constructed through actual social and material connections as well as through imputations, assumptions, and imagination.

**Formative and experiential dimensions of data spaces**

If the term data space is to be more than an evocative label, it needs specificity. As stated above, we suggest approaching data spaces theoretically as involving both formative dimensions (how they are enacted) and experiential dimensions (the aspects of human experience they evoke).

Beginning with the formative dimensions, we suggest thinking of data spaces as enacted through promises, work, and users. These dimensions are generic in the sense that they could be used for analyzing the construction of any data infrastructure, not just data spaces. It is our way of highlighting the political and social dynamics that are part of building material infrastructures. Promises come in many forms, as articulations of expectations, potential, commitment, or anticipation. They inform investment strategies, and thereby generate work, while work engages users. Again, their relationships can also be inverted: users produce work and articulate promises. We propose thinking of each of these elements without predefined assumptions about when agency stems from human or nonhuman actors. Computers can carry out work and become users themselves when networks set them up as actors in particular ways (Crawford, 2021; Rahman, 2021; Schuilenburg and Peeters, 2021; Seaver, 2018). Exploring promises also involves understanding why various stakeholders wish to integrate data sources, that is, their desires and ambitions (Tutton, 2017). The sociology of expectations inspires us to explore promises not as much for what they anticipate of the future but for what they do in the present (Borup et al., 2006; Brown, 2003; Brown and Michael, 2003; Pinel, 2022; Tutton, 2017; Vezyridis and Timmons, 2021).

Promises generate and legitimize investments, while investments in turn install work. Berg and Goorman (1999) famously formulated a “law of medical information” stating that: “The further information has to be able to circulate (i.e., the more different contexts it has to be usable in) the more work is required to disentangle the information from the context of its production” and encouraged us to explore: “who has to do this work and who reaps the benefits?” (p. 52). This question will be of central importance when the EHDS suggests data work that facilitates cross-border data sharing. With increased data integration, data work becomes ubiquitous (Bossen et al., 2019). Sometimes computers do part of the data work, but automation is never complete. Considerable human labor is involved in making data ready for machines (Fiske et al., 2019; Møller et al., 2020; Torenholt et al., 2020). This labor gives reason to question who or what is the user (Hyysalo et al., 2016; Oudshoorn and Pinch, 2003).

Whereas the formative dimensions—promises, work, and users—express how a data infrastructure comes into being, we also need a vocabulary for how the data exchange is experienced by people. The data space is the phenomenon that emerges as people live with data exchanges in their daily lives. To capture these experiences, we suggest focusing on four general dimensions of human experience: what is felt as right, true, present, and valuable. Attention to such experiential dimensions position people as something akin to “inhabitants” of data spaces. We need to acknowledge that people have emotionally charged experiences with data, even when they might not understand who has access to what, just as Ruckenstein (2023) has argued that we need to acknowledge that people have emotional reactions to algorithms they do not understand. Such experiences shape what data infrastructures do. They shape how people understand themselves (Amoore, 2020). Just like different people live very different lives even when occupying similar houses, people will inhabit similar data spaces differently—with different moral, epistemic, phenomenological, and value commitments. Habitation can involve making a living in data spaces without deliberately being part of building them, just like few people have built their own houses. This is important because people might not always be aware of the actual workings of the data infrastructures surrounding them and yet still be constrained by (or partake in shaping) what they do in practice.
If we begin with what is experienced as right, the anthropology of law has proposed that most human communities share a sense of legality, or something being normatively and legally sanctioned (Moore, 1978; Nader, 2002). Legality can be seen as a social process tied to practices and not just legal texts. Inspired by Lessig (2006), we might see regulation of digital technology as increasingly engrained in code. Software has come to shape social interaction in multiple ways (Kitchin and Dodge, 2011). What health professionals and patients orient themselves toward—in their understanding and experience of what is “right”—may come as much from their interaction with digital interfaces as from legal texts. Furthermore, they cannot be expected to agree about what is “right.” We can explore these experiences of what is “right” empirically as they unfold, and we need to do so to understand the normativity of data spaces.

Another key experiential dimension of data spaces is whether and how people come to think of data and data analyses as true. Traditionally, this has been a topic reserved for the philosophy of science, and here the idea that science provides a singular universal Truth (of the type once propagated by Auguste Comte) has largely been abandoned, along with oversimplified positivist propositions. Medical researchers rarely talk about data or data analyses as “true.” Rather, they focus on whether data are relevant, reliable, or valid, and whether their analysis can lead to results that are significant or can count as evidence. STS researchers and philosophers of science also talk about robust knowledge when results or understandings are shared by independent analyses or even by people in different positions (Oreskes, 2019). Not all data analyses of the same data are equally valid (Hand, 2020). Some analyses are even deliberately misleading, for economic or political reasons (Adams, 2016; Merry, 2016; McGoe, 2012). As an experiential dimension, however, people continue to orient themselves toward notions of truth, toward insights they trust, and we therefore suggest using “true” as a placeholder for the above terms indicating convictions of epistemological validity. We still acknowledge the point about data as faulty knowledge-devices elegantly formulated by Goldstein and Nost (2022) in relation to how people acquire environmental knowledge, namely that, “data and its infrastructures are not so much clear, objective windows as they are tinted lenses, where everything about them—from the tint to the frames—shape what and how we see, as well as what gets contested” (p. 12). Furthermore, we contend that in data spaces, what becomes established as “true” or “false” is not stable, and can be overruled by ideas about what is right, present, or valuable.

While legislators refer to the knowledge that data can be used to produce, they should not assume that people will automatically experience such knowledge as personally relevant, or present in their practical conduct of everyday life. Presence is another important experiential dimension of data spaces. It relates to phenomenological engagement with data. Data work on and through us. But how? It is not just a matter of cognition and analysis. To understand the experiential dimension of data spaces, we need to explore how people engage with data and data representations as bodies and as subjects. A few scholars have begun paying attention to the way data and data infrastructures can be experienced as “enchanting” (Smith, 2018) and “beautiful” (Halpern, 2014), and how data esthetics (color coding, graphs, and interactive maps) mediate power effects by privileging particular representations and silencing others (Kennedy and Hill, 2018; Ratner and Ruppert, 2019). Studies of self-trackers and patients using various wearables and data tools illustrate how people respond sensorially to seeing (Kristensen and Frigge, 2018) and sensing (Knox and Nafus, 2018) their own data. Data affect them, not as mere conveyers of information, but as “partners” they can interact with (Kristensen and Ruckenstein, 2018). Likely, health professionals and policy planners experience something similar when encountering data (Lupton and Maslen, 2017). Immaterial and diffuse goals become material when, for example, red turns to green in a cell phone app or when a pedometer makes a sound. With data spaces as a conceptual framework, analysts can attune to how people phenomenologically engage with data and data representations—not always by establishing their message as true, but by making them feel present in their own lives and bodies.

Finally, when moving in data spaces, people make assessments of what is valuable. Data carry multiple forms of value (Asdal et al., 2021; Leonelli, 2016: 64). The political interest in European cross-border data infrastructures reflects economic perceptions of data as assets in a digital economy (Birch et al., 2021; Sadowski, 2019). As Sharon argues, however, the forms of value at stake for different users cannot be reduced to economic valuation alone, and the competing registers of value can generate friction (Sharon and Zandbergen, 2016; Sharon, 2021). The value of data—in which form and for whom—depends on the context and the purpose for which they are used (Fiske et al., 2022). However, context dependency and friction between the purposes of different users are ignored when the Commission, for example, writes, “The European Health data space will benefit individuals, health professionals, healthcare providers, researchers, regulators, and policy-makers” (European Commission, 2022a: 12). Such formulations suggest that data just have to be set free, then everybody will benefit.

In sum, we suggest conceiving of data spaces as formed by promises, work, and users and experienced in at least four dimensions relating to what is right, true, present, and valuable. The dimensions are summarized in Table 1. It is of course possible to think of other dimensions. Esthetics, for example, might be influencing each of the four (Kennedy...
and Hills, 2018), just as trust and legitimacy might be a form of cross-cutting outcome. Still, these initial conceptual delineations are meant to help articulate data spaces as objects of governance in a new and more comprehensive way than what normally transpires from policy documents.

Equipped with this conceptual understanding of data spaces, we now turn to the way the European Commission suggests establishing an EHDS. We describe what policymakers claim to be their ambitions and relate each ambition to the formative and experiential dimensions of the data spaces they are trying to build. While the EHDS is still only in its planning phase, it is interesting to read the proposal as exemplifying ideas about a governmental object devoid of the very social thickness we have just argued that data spaces entail.

The European Health Data Space: Three ambitions in tension

To understand the ambitions of the European Commission, we focus on the text of the Proposal for a Regulation on the European Health Data Space from May 3, 2022 (European Commission, 2022b). We are aware that the shaping of cross-European infrastructures for health data is embedded within a wider set of ongoing policy initiatives aimed at establishing a competitive European data economy, including the General Data Protection Regulation (GDPR), the Data Governance Act, the Digital Markets Act, and the Digital Services Act, as well as strengthened public health surveillance through the European Centre for Disease Prevention and Control. Our focus lies, however, on the EHDS proposal, since this is the most comprehensive piece of legislation aimed directly, explicitly, and exclusively at the establishment of cross-border health data infrastructures. A regulation like the EHDS, unlike a directive, will have immediate legal effect in all member states. Although the EU has limited competence in healthcare, the EU has been finding indirect ways to exercise its policymaking powers in this area (de Ruijter, 2019). One way the EU can justify intervention in the field of health is when all countries face challenges in need of shared action. It is interesting how data integration has become one such concern.

Figure 1 displays how the EU Commission originally outlined the overall goal of EHDS. Across the various reports and press releases of the EHDS, the specific formulations of goals have already changed significantly, and the goals are bound to change again before the actual legislation is passed—just as they will keep changing when the EHDS translates into a multiplicity of practices. Still, three ambitions appear consistently in the policy papers: (1) to enhance individual citizens’ control over their own health data, (2) to facilitate reuse of health data across national boundaries for secondary purposes, and (3) to harmonize markets for digital health products, including electronic health records (European Commission, 2022b: 1). While the aim of harmonization is expected to support that of reuse, these latter two aims do not easily align with the first aim of enhanced individual control. In fact, they could easily conflict. Furthermore, throughout the proposal, it remains unclear what constitutes use and reuse. In research, this has never been easy to delineate (Dallmeier-Tiessen et al., 2019), but the proposal in a sense now places clinical and other purposes on equal footing, whereby potential conflict between data needs is underplayed. Policymakers frame data spaces as a matter of wired connections—a technical sense of space—that just need safeguards and legal demands to work well. However, such a reduction to technicalities misconstrues the governmental object targeted by the EHDS. There is a need for understanding much more comprehensively the social and political dynamics that bring data spaces into being as well as the experiential dimensions of living with data exchanges. We now turn to what the EU Commission concretely suggests, beginning with a closer look at the ambition of enhancing individual control.

The EHDS as empowerment of citizens

In the proposal, article 3 states:

Natural persons shall have the right to access their personal electronic health data processed in the context of primary use of electronic health data, immediately, free of charge and in an easily readable, consolidated and accessible form.

Besides access, individuals shall also have a right to insert data into the electronic health records (EHRs) kept by their healthcare providers:

Natural persons may insert their electronic health data in their own EHR or in that of natural persons whose health information they can access, through electronic health data access services or applications linked to these services. (§3(6))

If users in the data space embrace the meaning of these articles in their most literal sense, their perception of what is right will be influenced, as they will expect quite far-reaching powers concerning data access and control. The mere ability to upload and access data will significantly change everyday practices in the health services for many European citizens. In some EU countries, including Denmark and Sweden, citizens have had online access to their health records for years, but never to all the data collected in conjunction with their healthcare. In Denmark, for example, patients have never had direct access to genetic data or imaging data (which they are not expected to understand), and following a major political conflict in Denmark in 2014, patients no longer have access to records from general practice (GP) (Langhoff et al., 2018). GPs experienced upset patients
who had accessed diagnostic information they interpreted very differently from the doctor’s intentions (Wadmann and Hoeyer, 2018). While the data were not acknowledged by patients as true, they had become present in patients’ understanding of the care they received and their reactions challenged the stability of the data space.

These challenges illustrate that the value associated with empowerment through online access may erode values associated with clinical uses of health information. Value assessments often conflict. In data-intensive health systems, many clinical conversations now start with questions about data rather than beginning with clinical symptoms (Kristensen et al., 2021; Petersson and Backman, 2021). Externally imposed data standards for the sake of facilitating nonclinical uses will influence the clinical uses and the conventions of patient care. In both Denmark and Sweden, studies have shown that patient access to data affects registration practices (Petersson, 2019). Changing registration practices furthermore means that the validity of some data can be undermined, which also holds implications for research and administrative uses of data. Questions also arise surrounding the requirement of “easily readable” formats for electronic health records, as “readability” does not always mean the same for patients as for clinicians, researchers, administrators, and algorithms.

The potential value conflict between clinical and other uses is also clearly present in the tension between the stated ambition of patient empowerment and the mandatory requirement of the EHDS proposal to share health data for multiple purposes—though not acknowledged as a conflict by the Commission. While patients will have access to see and upload their own data, they have much less control over how data are used and by whom. This shift may not be compatible with the data subjects’ rights established in the GDPR (European Data Protection Board (EDPB) and European Data Protection Supervisor (EDPS), 2022), which may not be circumvented simply by arguing that data are anonymized as nonpersonal because the boundaries of this categorization are increasingly blurry (Marelli et al., 2023). Article 33(5) suggests letting data access bodies overrule existing demands for informed consent. Furthermore, the current proposal specifies that patients do not even have a right to know how their data are being used, by whom, or for what (article 38). How would patients experience such a legal setup? Which data uses will the parts of the population with limited trust in the authorities begin to imagine are taking place? Is it reasonable to assume that a legislative change will be enough for people to experience data exchanges as “right”?

Furthermore, the explanatory memorandum accompanying the proposal contains an interesting formulation of the moral and legal obligation to share: the regulation is intended to “establish mechanisms for data altruism in the health sector” (p. 2). Here the commission imbues the word “altruism” with a form of goodhearted sharing-by-default, even when patients have no way of finding out what they share, to what aim(s) they share, or even that they share. Again, we will have to explore how this will affect patients’ experience of what is “right” and “valuable.” The EHDS proposal impinges not only on the terrain of wires and entitlements, but also that of social obligations and relations among citizens and between citizens and institutions. In the Nordic countries, the construction of infrastructures that facilitate research uses of clinical data has taken decades: they are engrained in welfare state policies and depend on high degrees of trust (Svendsen and Svendsen, 2015). It is naïve to assume that these social aspects of data access can be installed overnight by legal decree.

**EHDS as a means of establishing authority over data to enable reuse**

To facilitate cross-border data sharing, the EU Commission envisages not just mandating the building of new infrastructures,
but also requiring the establishment of organizations in each member state, called health data access bodies, responsible for ensuring data overviews, accumulating metadata, and providing access. If a company in one country approaches a national health data access body, that organization is supposed to ensure access to the desired data in the entire EU. The EHDS proposal thereby proposes to establish a new type of authority. If implemented, this will install new forms of work on part of the agencies and health professionals who need to provide the data (Green et al., 2023). Health professionals will have limited control over the data they produce (Wadmann et al., 2023).

Which data are healthcare providers going to be obliged to share with EHDS? Article 33 lists the “Minimum categories of electronic data for secondary use” as:

(a) electronic health records; (b) data impacting on health, including social, environmental behavioural determinants of health; (c) relevant pathogen genomic data, impacting on human health; (d) health-related administrative data, including claims and reimbursement data; (e) human genetic, genomic and proteomic data; (f) person generated electronic health data, including medical devices, wellness applications or other digital health applications; (g) identification data related to health professionals involved in the treatment of a natural person; (h) population wide health data registries (public health registries); (i) electronic health data from medical registries for specific diseases; (j) electronic health data from clinical trials; (k) electronic health data from medical devices and from registries for medicinal products and medical devices; (l) research cohorts, questionnaires and surveys related to health; (m) electronic health data from biobanks and dedicated databases; (n) electronic data related to insurance status, professional status, education, lifestyle, wellness and behaviour data relevant to health; (o) electronic health data containing various improvements such as correction, annotation, enrichment received by the data holder following a processing based on a data permit.

The list is admittedly overwhelming and long; some readers might have decided to not really read the list and simply skimmed through it. This is precisely why we cite it in full here. We have participated in meetings where people with extensive experience in health data integration have looked at this list and heard them laugh because it appears utterly unrealistic. Still, one civil servant conveyed to Klaus Hoeyer that a preproposal had been circulated which was even more far-reaching, which had made this official version appear less “extreme.” It now, as the text says, pertains to just “minimum categories.”

In addition to the primary data in the list above, national health data access bodies are to produce metadata describing the quality, completeness, uniqueness, accuracy, validity, timeliness, and consistency of the data (article 56). This may sound nice to have, and indeed metadata are needed to make sense of the primary data, but they are anything but easy to produce. This will demand extensive labor. Work, as a formative dimension, is essential to understand thoroughly when envisaging a data space: data are never just “there”—they are made to be in particular databases in particular formats through work (Leonelli, 2016; Bossen et al., 2019). The budget calculation annexed to the proposal, however, says that metadata are not supposed to imply extra resources. This is an utterly unrealistic understanding of costs.

European Commission lawmakers think of secondary use of health data as legitimate (“right”) because the data are only intended to be analyzed on platforms that do not contain individual identification. It cannot be taken for granted, however, that patients and health professionals trust such anonymization. Moreover, patients and health professionals may not agree with the new purposes to which data are assigned. They can have other ideas about the “right” purposes for data reuse. Furthermore, when people are uncertain about where data circulate, they typically start guessing (Ruckenstein, 2023). The data spaces in which people move will not be confined to the actual exchange of data: the experiential space will comprise assumptions, guesses, and imputations. The policy promises of a seamless data flow across boundaries and purposes further invite such guesswork.

Which purposes are described in the proposal as legitimate—and who are to be able to access data? Again, the list is quite extensive. According to article 34, public sector bodies are to use the data for:

(a) activities for reasons of public interest in the area of public and occupational health, such as protection against serious cross-border threats to health, public health surveillance or ensuring high levels of quality and safety of healthcare and of medicinal products or medical devices; (b) to support public sector bodies or Union institutions, agencies and bodies including regulatory authorities, in the health or care sector to carry out their tasks defined in their mandates; and (c) to produce national, multi-national and Union level official statistics related to health or care sectors.

It may already seem broad, but article 45 furthermore states that “Any natural or legal person may submit a data access application” (our emphasis) for the following purposes (also listed in article 34):

(d) education or teaching activities in health or care sectors; (e) scientific research related to health or care sectors; (f) development and innovation activities for products or services contributing to public health or social security, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices; (g) training, testing and evaluating of algorithms, including in medical
devices, AI systems and digital health applications, contributing to the public health or social security, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices; (h) providing personalised healthcare consisting in assessing, maintaining or restoring the state of health of natural persons, based on the health data of other natural persons.

Again, the list might simply be too long for the reader to bother reading thoroughly, and again we present it in full to illustrate how very few limits constrain the imagined purposes data are to serve. Each purpose carries a valuation, and it will be interesting to follow empirically how health professionals and patients perceive these diverse values. What the data space becomes in practice will depend partly on how potential value conflicts are resolved. Not least for this reason, the Commission needs to rethink the governmental object.

Article 35 specifies purposes that shall be prohibited. Electronic health data may not be used to make decisions that may be “detrimental to a natural person.” To qualify as decisions, “they must produce legal effects or similarly significantly affect those natural persons.” Article 35 also suggests prohibiting data use to exclude persons or groups of persons “from the benefit of an insurance contract or to modify their contributions and insurance premiums.” Such a rule would significantly alter the insurance market. It is not only a legal change: it will have economic and social implications. Furthermore, data may not be used for advertising or marketing activities geared towards health professionals, organizations in health, or citizens (in the text named “natural persons”), nor for “developing products or services that may harm individuals and societies at large.” The examples mentioned include “illicit drugs, alcoholic beverages, tobacco products, or goods or services which are designed or modified in such a way that they contravene public order or morality.” What counts as contravening “public order or morality” remains an open question, and potentially a very controversial value issue in a union known for containing diverse and conflicting views on sexual minorities, abortion, and reproductive technologies. Still, there is something very sympathetic—and potentially game-changing—about prohibiting data use that can harm the individual or promote marketing activities.

Different experiential dimensions of living in such data spaces may overrule or conflict with one another. For instance, the “truth” of data is not unequivocal. Data such as diagnostic codes can be found to be outright “wrong” in a research setting—that is, a patient who did not have cancer gets misrepresented by such a diagnostic code. However, while the code may be “wrong” in terms of a medical diagnosis, it may remain “correct” historical documentation of a hospital error, and therefore one that cannot be erased. Furthermore, the absence of data may contribute to impreciseness without any data being inherently “wrong” (Hand, 2020). As datasets and data processing are used further and further away from the origin of data production, misunderstandings of data are likely to increase (Loukissas, 2019). These issues of validity become all the more important when EHDS data are to be used for training AI (Hildebrandt, 2023), and if people are met with AI trained on EHDS data, it will shape their experiences of care in profound ways. Once industry also begins using access to data from electronic health records to make claims about drug safety, efficacy, or value, the truth value of data may become even more contested.

The EHDS as a way of harmonizing markets for digital health products

The third key policy ambition for the EHDS is harmonizing markets for digital health products, especially electronic health records. This ambition aims toward standard setting. While this can be interpreted as merely a matter of supporting the aim of data repurposing by making everybody work in interoperable formats, we suggest there is more at play. Standard setting is political to its core, with wide-ranging implications for reporting and use of data (Busch, 2011; Dunn, 2005; Grommé and Ruppert, 2020; Timmermans and Berg, 1998). It is by establishing standards many of the value conflicts above are settled and users are prioritized.

The EHDS proposal suggests that the Commission shall be granted the power to specify a “European electronic health record exchange format” (article 6), that manufacturers of electronic record systems shall ensure conformity with (article 17) and ensure prior approval of before offered on the European market (article 18). Commenting on this demand, one industry representative said at a meeting to Klaus Hoeyer: “Finally, clinical data will also become FAIR.” With FAIR, he meant Findable, Accessible, Interoperable, and Reusable, in the same manner as research data are supposed to be FAIR in the Union. There are, however, profound differences between research data from standardized clinical trials and from clinical practices in 27 different healthcare systems. A shared data format does not mean that people register in similar ways. It is an old insight that even shared coding manuals are used differently depending on the structure of the healthcare systems in which they operate (Winthereik, 2003). Data are entangled in specific uses, not stable entities with unambiguous meanings (Thyrlstrup et al., 2022).

If legislators learn to think of data spaces as composed of formative and experiential dimensions—that is, as socio-technical constructs emerging through practices—and not just as wires and legal entitlements, they might also be better equipped to explore what is already known about how standard setting operates in practice. Health professionals can get very frustrated when asked to use digital
tools that they find ill-suited for clinical purposes (Hunt et al., 2017; Gawande, 2018), or when they learn that data are used for purposes they deem illegitimate (Langhoff et al., 2018). They begin using the tools in unexpected ways: they change registration practices (Hoeyer & Wadmann, 2020). Because data spaces come into being through users and work, and because users are affected by the data spaces they inhabit, legislators must engage a much broader understanding of what they are seeking to establish than what is now called "standards."

Interestingly, the Commission appears to be aware of potential resistance or, at least, limited interest among healthcare professionals in the suggested changes. In the explanatory memorandum accompanying the EHDS proposal, they write that the results of evaluations and consultations show that the "The evaluation of the eHealth provisions under the [Cross Border Health Care Directive, 2011/24/EU] concluded that its effectiveness and efficiency has been rather limited and that this was due to the voluntary nature of the eHealth Network actions" (p. 9). They also note that—

...some Member States set up different bodies to deal with the subject and participated in the joint action Towards a European Health Data Space (TEHDaS). However, neither this joint action, nor the numerous funds provided by the Commission (...) to support the secondary use of electronic health data have been sufficiently implemented. (p. 9)

In working to prepare the proposal, the lawmakers have also commissioned a study of health data reuse that suggests that clinical data rarely are repurposed, even when available (European Commission and Directorate-General for Research and Innovation, 2021). However, none of these studies seem to influence the high expectations of the Commission. Instead, the current lack of use is deployed to motivate new investments, as when the Commission writes: “today’s EU health sector is rich in data, but poor in making it work for people and science” (European Commission, 2022a: 1). Perhaps the lawmakers think they can overcome resistance and lack of interest by enforcing their vision with digital means, as was the case with the HITECH Act in the USA, which lead to digitization of health records (Greene, 2022: 245). However, if the Commission expects digital technologies to be able to crush resistance, they are misunderstanding how a data space is enacted. There is no way to entirely control users, and no way to suppress the variance in experiences. For the EHDS to succeed, the Commission must rethink the governmental object.

Conclusion

With the EHDS, the European Commission has initiated a major transformation of European healthcare. It has done so, however, based on an inadequate understanding of what data integration implies. As seen in other areas of EU legislation on digital integration (Bellanova and Duez, 2012), the proposal’s text propagates a view of data integration as merely legal (defining entitlements and realms of authority) and technical (connecting wires and establishing standards) issues. To help legislators, as well as the practitioners who will deal with the practical implications, we have proposed a more adequate conceptualization of “data space.” We suggest that the governmental object of the EHDS will emerge from the interplay of at least three formative (promises, work, and users), and four experiential (right, true, present, and valuable) dimensions. These formative enactments and experiential dimensions interact. Work, for example, involves experiences. These experiences are likely to influence who becomes users (or nonusers). Experiences of what is right, true, present, and valuable possibly shape promises and how they are received by different groups of people. When infrastructures achieve their aims of seamless data integration, Star reminded us, they tend to become invisible (Star, 1991). They are no longer “experienced” as objects. Likely, many citizens will never “experience” the final product of initiatives like the EHDS in any conscious manner, but they will nevertheless inhabit the resulting data spaces and may experience both what integrated data infrastructures do and fail to do.

We thus present a way of thinking of data spaces as sociotechnical enactments. This is important because a data space cannot be legally and technically controlled: it will always exceed what engineers build, and it will continue to change and surface new experiences. Therefore, we also want this conceptualization to pave new paths of research, where social studies of data dynamics involve continuous monitoring and learning from practice. A theoretically informed understanding of data spaces will, we believe, create more realistic ambitions. Legislators will benefit from thinking about cross-border data integration as emerging out of not just legal and technical design but through promises, users, and work—that is, through enactments. These enactments will engage different experiences in reflection of moral appraisal (what people see as right), epistemological convictions (what they believe to be true), phenomenological engagement (what becomes present to them), and their perceptions of worth (what they deem valuable). Experiential dimensions are shaped by not only actual data exchanges, but also imputations, assumptions, and imagined intentions. As such, they are uncontrollable.

To monitor what their proposed standards do, the EU and the member states will need to attend to how these elements enact the EHDS in a multiplicity of ways. Data standards shape work, but they are also shaped by work. In very different healthcare systems offering very different services, the same standards will involve diverse types of work by distinct users and thereby create different types
of data. From this perspective, the promise of comparable, standardized, and interoperable data is bound to fail or, rather, produce something significantly other than what the promises of the EHDS suggest. The administrators, practitioners, and researchers, who are to establish and use the EHDS, therefore need more realistic ways of thinking about the governmental object produced by the EHDS. More realistic conceptions of data spaces can help ensure better plans for monitoring progress. They may help build more socially sustainable infrastructures better fitted to the needs and desires of their users—infrastructures that are accessible to both majority and minority populations.

Though our critique of the EHDS proposal can be read as an attempt to expose an apparent naïveté among policymakers, we do acknowledge that the proposal text simply exemplifies a particular powerful discourse—what can be said on that topic in those circles at this moment in time—not the thinking of individual lawmakers. Many policymakers might very well be aware of some of the limitations we describe. Still, as such a discourse comes to shape the tools available for the health services, we need to create other discourses that facilitate different types of reflections in order to gradually build more socially robust data infrastructures.

There is an inherent tension in the proposal between empowerment of citizens and loss of control over data. This tension also marks other data integration initiatives in healthcare (Hoeyer, 2023). This tension is not merely legal - it is also experiential. Infrastructures such as EHDS are likely to make data exchanges in healthcare subject to the same type of guesswork and inference that emerges among people when they are surprised by a particular piece of advertising on Facebook and Google and begin wondering whether they have been tracked or their phones have been eavesdropping (Ruckenstein, 2023). Yet the stakes are likely much higher for an individual when related to their personal health, as opposed to commercial advertising or consumer recommender systems. These stakes may impact not only people’s relationships with health professionals and healthcare systems, but also how they think of themselves and their health as data points that can flow and be utilized across borders and users. While we are building data infrastructures, they are building us. If Geertz once suggested that “man is suspended in webs of significance he himself has spun” (Geertz, 2017: 5), we can now perceive how everybody is beginning to live in the webs of data they themselves have spun. It is a life suspended in data space.

Acknowledgements
The authors would like to thank the editors and three generous reviewers, as well as Ine Van Hoyweghen and Patrik Hummel, for extremely helpful and insightful comments on earlier versions. Hoeyer presented an earlier version of this paper as the Andrew J Webster memorial talk in York, September 7, 2022, and would like to thank the organizers and the audience for comments.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement number 101096999). Alexandra Middleton would also like to acknowledge funding from the European Research Council (ERC) (grant number 949050). Clémence Pinel would also like to acknowledge funding from the Carlsberg Foundation (grant number CF17-0016).

ORCID iDs
Klaus Hoeyer https://orcid.org/0000-0002-2780-4784
Sara Green https://orcid.org/0000-0001-7011-7202
Andrea Martani https://orcid.org/0000-0003-2113-1002
Alexandra Middleton https://orcid.org/0000-0002-1967-5994
Clémence Pinel https://orcid.org/0000-0003-3874-0373

Notes
1. Research of this type is not subject to ethics approval in Denmark. We have participated in meetings and webinars about the EHDS organized by interested parties as well as official stakeholders and we have been invited for informal meetings where selected stakeholders have asked for our input on the proposal and in this process learned about their views and understandings. The discussions have helped us better understand the proposed legislation. Furthermore, Klaus Hoeyer has been granted access to observe the meetings of one of the groups charged with doing an EHDS test case for research uses. While these observations have sharpened our understanding, and sometimes delivered illuminating anecdotes, they are not used as the foundation for the claims we make. Rather, we use published research to challenge the understandings transpiring in the proposal text.
3. In health geography, a “space” is typically thought of as a Euclidean basis (3 dimensions) for “place” as a semantic understanding of that space (Thomson, 2021), while in social anthropology it is typically the other way around (Green, Harvey, and Knox, 2005). We here adhere to the anthropological custom of seeing space as socially constituted and enacted.

References


